



---

Year: 2021

---

## Spatial Distribution of Private Gene Mutations in Clear Cell Renal Cell Carcinoma

Moore, Ariane L ; Batavia, Aashil A ; Kuipers, Jack ; Singer, Jochen ; Burcklen, Elodie ; Schraml, Peter ; Beisel, Christian ; Moch, Holger ; Beerenwinkel, Niko

**Abstract:** Intra-tumour heterogeneity is the molecular hallmark of renal cancer, and the molecular tumour composition determines the treatment outcome of renal cancer patients. In renal cancer tumourigenesis, in general, different tumour clones evolve over time. We analysed intra-tumour heterogeneity and subclonal mutation patterns in 178 tumour samples obtained from 89 clear cell renal cell carcinoma patients. In an initial discovery phase, whole-exome and transcriptome sequencing data from paired tumour biopsies from 16 ccRCC patients were used to design a gene panel for follow-up analysis. In this second phase, 826 selected genes were targeted at deep coverage in an extended cohort of 89 patients for a detailed analysis of tumour heterogeneity. On average, we found 22 mutations per patient. Pairwise comparison of the two biopsies from the same tumour revealed that on average, 62% of the mutations in a patient were detected in one of the two samples. In addition to commonly mutated genes (VHL, PBRM1, SETD2 and BAP1), frequent subclonal mutations with low variant allele frequency (<10%) were observed in TP53 and in mucin coding genes MUC6, MUC16, and MUC3A. Of the 89 ccRCC tumours, 87 ( 98%) harboured private mutations, occurring in only one of the paired tumour samples. Clonally exclusive pathway pairs were identified using the WES data set from 16 ccRCC patients. Our findings imply that shared and private mutations significantly contribute to the complexity of differential gene expression and pathway interaction and might explain the clonal evolution of different molecular renal cancer subgroups. Multi-regional sequencing is central for the identification of subclones within ccRCC.

DOI: <https://doi.org/10.3390/cancers13092163>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-203642>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Moore, Ariane L; Batavia, Aashil A; Kuipers, Jack; Singer, Jochen; Burcklen, Elodie; Schraml, Peter; Beisel, Christian; Moch, Holger; Beerenwinkel, Niko (2021). Spatial Distribution of Private Gene Mutations in Clear Cell Renal Cell Carcinoma. *Cancers*, 13(9):18.

DOI: <https://doi.org/10.3390/cancers13092163>

## Article

# Spatial Distribution of Private Gene Mutations in Clear Cell Renal Cell Carcinoma

Ariane L. Moore <sup>1,2,†</sup> , Aashil A. Batavia <sup>1,2,3,†</sup>, Jack Kuipers <sup>1,2</sup>, Jochen Singer <sup>1,2</sup>, Elodie Burcklen <sup>1</sup>, Peter Schraml <sup>3</sup>, Christian Beisel <sup>1</sup>, Holger Moch <sup>3,\*</sup>  and Niko Beerenwinkel <sup>1,2,\*</sup> 

<sup>1</sup> Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland; ariane.moore@bsse.ethz.ch (A.L.M.); aashil.batavia@usz.ch (A.A.B.); jack.kuipers@bsse.ethz.ch (J.K.); jochen.singer@bsse.ethz.ch (J.S.); elodie.burcklen@bsse.ethz.ch (E.B.); christian.beisel@bsse.ethz.ch (C.B.)

<sup>2</sup> SIB Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland

<sup>3</sup> Department of Pathology and Molecular Pathology, University of Zurich and University Hospital Zurich, Schmelzbergstrasse 12, 8091 Zurich, Switzerland; Peter.Schraml@usz.ch

\* Correspondence: holger.moch@usz.ch (H.M.); niko.beerenwinkel@bsse.ethz.ch (N.B.)

† Shared first author.

**Simple Summary:** Tumours consist of multiple groups of similar cells resulting from differing evolutionary trajectories, i.e., subclones. These subclones are prevalent in clear cell renal cell carcinoma (ccRCC). The aim of this study is to determine how similar or dissimilar the subclones in 89 ccRCC tumours are from one another regarding their gene mutations and expression profiles, i.e., the extent of intra-tumour heterogeneity. The implications of these alterations with respect to signalling pathways is also assessed. Deep sequencing allows for the identification of mutations with low-allele frequencies, providing a more comprehensive view of the heterogeneity present in the tumours. With an average of 62% of mutations having been identified in only one of the two biopsies, some of which in turn are found to impact gene expression, the complex makeup of ccRCC tumours is evident, and this can drastically influence treatment outcome.

**Abstract:** Intra-tumour heterogeneity is the molecular hallmark of renal cancer, and the molecular tumour composition determines the treatment outcome of renal cancer patients. In renal cancer tumourigenesis, in general, different tumour clones evolve over time. We analysed intra-tumour heterogeneity and subclonal mutation patterns in 178 tumour samples obtained from 89 clear cell renal cell carcinoma patients. In an initial discovery phase, whole-exome and transcriptome sequencing data from paired tumour biopsies from 16 ccRCC patients were used to design a gene panel for follow-up analysis. In this second phase, 826 selected genes were targeted at deep coverage in an extended cohort of 89 patients for a detailed analysis of tumour heterogeneity. On average, we found 22 mutations per patient. Pairwise comparison of the two biopsies from the same tumour revealed that on average, 62% of the mutations in a patient were detected in one of the two samples. In addition to commonly mutated genes (*VHL*, *PBRM1*, *SETD2* and *BAP1*), frequent subclonal mutations with low variant allele frequency (<10%) were observed in *TP53* and in mucin coding genes *MUC6*, *MUC16*, and *MUC3A*. Of the 89 ccRCC tumours, 87 (~98%) harboured private mutations, occurring in only one of the paired tumour samples. Clonally exclusive pathway pairs were identified using the WES data set from 16 ccRCC patients. Our findings imply that shared and private mutations significantly contribute to the complexity of differential gene expression and pathway interaction and might explain the clonal evolution of different molecular renal cancer subgroups. Multi-regional sequencing is central for the identification of subclones within ccRCC.

**Keywords:** intra-tumour heterogeneity; private mutations; clonal exclusivity



**Citation:** Moore, A.L.; Batavia, A.A.; Kuipers, J.; Singer, J.; Burcklen, E.; Schraml, P.; Beisel, C.; Moch, H.; Beerenwinkel, N. Spatial Distribution of Private Gene Mutations in Clear Cell Renal Cell Carcinoma. *Cancers* **2021**, *13*, 2163. <https://doi.org/10.3390/cancers13092163>

Academic Editor: Simon J. Furney

Received: 1 March 2021

Accepted: 27 April 2021

Published: 30 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

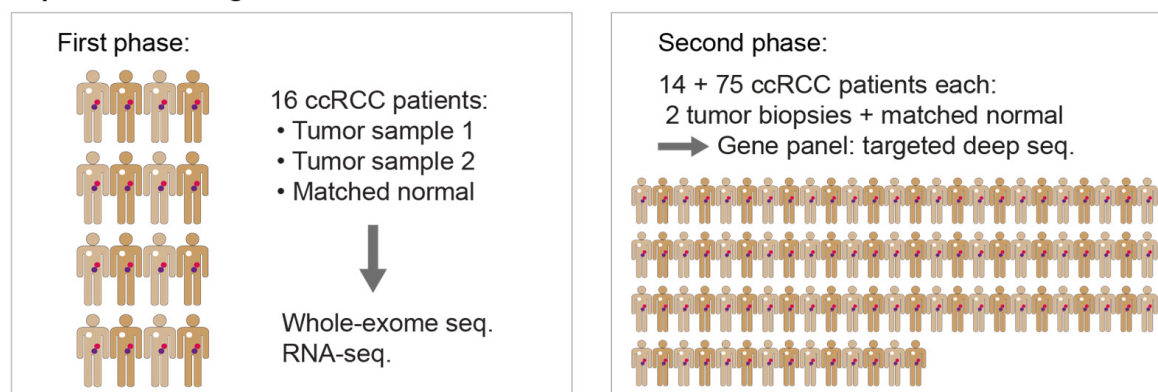
## 1. Introduction

Tumours consist of genetically and phenotypically distinct cancer cell populations that evolve over time through a process that involves mutation and selection [1]. The presence of intra-tumour heterogeneity is well founded in renal cell carcinoma (RCC) with multiple subclones in both the primary tumour and paired metastasis [2–6]. Gerlinger et al. assessed the heterogeneity within 10 renal carcinomas, applying multi-regional sequencing. A large degree of intra-tumour heterogeneity with respect to both somatic mutations and somatic copy number variations was observed in all 10 tumours, with 75% of driver events found to be subclonal [4,7]. Martinez et al. further showed in eight RCC that the diversity within tumours is in some cases as high as the diversity between patients [8]. Therefore, the number of somatic mutations may be undervalued when taking a single biopsy from a solid tumour with only a subset of clones being present in the metastasis. This work laid the foundation for the TracerX consortiums analysis of 101 RCCs with 1206 multi-regional samples. When assessing the metastasis of these tumours, it was found that the majority of the diversity accumulated in the primary tumour. It is within these primary tumours where metastasis-competent subclones undergo selection [9]. The identification of subclonal mutations is clinically relevant following observations that even low-frequency clones can carry markers of prognosis and drive the process of metastasis [4]. Independent of genetic heterogeneity, Okegawa et al. suggested that intra-tumour heterogeneity also presents itself in the form of metabolic differences between tumour cells, further demonstrating the complexity present within renal cancer [10].

Current treatment strategies for RCC include antiangiogenic and immune therapies, the latter being effective in only a subset of cases [11–13]. Very recently, Motzer et al. performed integrative multi-omics analyses of 823 renal carcinomas from a randomised phase III clinical trial (IMotion 151) and identified seven robust molecular subtypes [14]. These molecular subgroups were associated with differential clinical outcomes following a combination of an anti-angiogenesis agent (AA; bevacizumab, anti-VEGF) and an immune checkpoint inhibitor (ICI; atezolizumab or anti-PD-L1) versus a VEGF receptor tyrosine kinase inhibitor (sunitinib).

Here, we investigate the subclonal mutation composition of clear cell renal cell carcinoma in two steps (Figure 1). In the initial discovery phase, we analysed two spatially separated biopsies and a matched normal sample from each of 16 ccRCC patients to provide an overview of the diversity and to inform the selection of genes for the second in-depth follow-up analysis. In the second phase, we used the constructed gene panel to sequence 826 genes at high coverage in 178 paired tumour samples and 89 matched normal samples from 89 ccRCC patients. We found frequent subclonal mutations in *TP53* and in mucin coding genes *MUC6*, *MUC16*, and *MUC3A*. Further, we tested for clonal exclusivity to identify combinations of signalling pathways that co-exist in the same tumour but in different tumour cell clones.

## Experimental design



**Figure 1.** Experimental design. The first phase includes 16 clear cell renal cell carcinoma (ccRCC) patients of which two spatially separated biopsies from the primary tumour and a matched normal sample were collected. Whole-exome sequencing and transcriptome sequencing was performed and the detected mutations informed the selection of genes for the panel of the second phase. The second phase includes an extended cohort of patients and the selected genes were targeted with higher coverage. From a total of 89 patients, we analysed two spatially separated tumour biopsies and a matched normal sample per patient. Fourteen of the patients in this panel data set were also among the 16 from the first phase.

## 2. Results

The assessment of whole-exome and transcriptomic sequencing data from paired tumour biopsies along with a matched normal biopsy from 16 ccRCC patients revealed an average of 40% of mutations were private and 31% of genes were differentially expressed in only a single biopsy. A gene panel was produced consisting of 826 genes and targeted for sequencing at deep coverage in a larger cohort of 89 patients, each with paired tumour biopsies and a matched normal biopsy. We found that the mutational frequencies of the most commonly aberrated genes in ccRCC found in our cohort are comparable to those expected given data from previous large cohort studies. With the identification of low-frequency mutations following deep sequencing, the average number of private mutations increased to 62%. After the assignment of mutations to clones using the tool Cloe, enrichment and pathway-level clonal exclusivity analysis was applied to identify clonally exclusive pathway pairs.

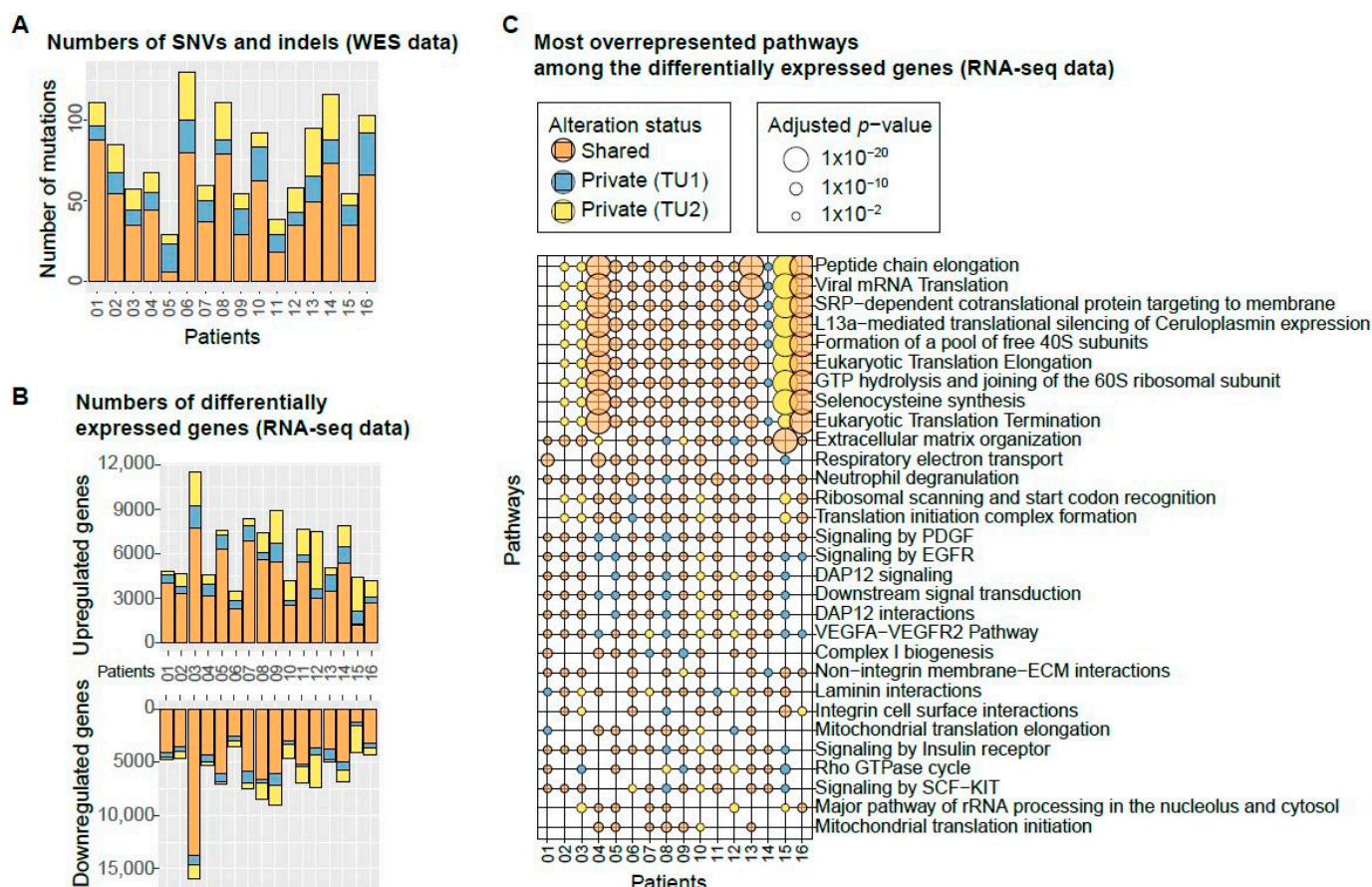
### 2.1. Genetic and Transcriptomic Diversity in 16 Ccrcc Patients

The coverage of the whole-exome sequencing (WES) data was on average 85x, and mutation calling (see Methods) identified between 29 and 130 single-nucleotide variants (SNVs), insertions, and deletions (indels) per patient (Figure 2A). The fraction of mutations that was only detected in one of the two biopsies from the same tumour was on average 40%, which indicates high levels of intra-tumour genetic diversity. These mutations are referred to as private, whereas mutations detected in both tumour samples of a patient are known as shared.

From the RNA-sequencing (RNA-seq) data, the differentially expressed genes were called by comparing each tumour sample to its paired normal using both single and paired-end data (see Methods). We found an average of 6364 genes per patient to be upregulated and 6598 genes downregulated (Figure 2B) with an average of 31% of differentially expressed genes being detected only in one of the two biopsies. Pathway overrepresentation analysis was performed with the set of differentially expressed genes using the Reactome pathway database [15]. Among the most overrepresented pathways are many pathways related to translation, signal transduction and growth factors (Figure 2C). The signalling pathways involving the growth factors PDGF, VEGF, SCF, or the growth factor receptor EGFR are deregulated in many patients. Of note, the vascular endothelial growth factor A (VEGFA), important for angiogenesis, cell growth, and survival [11] is upregulated in



all patients of this data set. The most overrepresented pathways related to translation are highly overrepresented in patients 4, 15, and 16. They are enriched only privately in one tumour sample of patients 2, 3, 14 and 15 each, indicating that these deregulated processes are subclonal in these tumours. Assessment of the similarity between samples as measured by the Euclidean distance showed a clear separation between tumour and normal samples, as expected. Within patient 3, where a large number of differentially expressed genes were identified, and patient 15, where multiple privately enriched pathways were found, TU1 and TU2 samples are more distant in comparison to the other patients which cluster according to their sequencing method (Supplementary Figure S1 and Additional File S2).



**Figure 2.** Genetic and transcriptomic diversity in 16 patients. **(A)** Number of shared (orange) and private (yellow, blue) mutations in the WES data set. A shared alteration was detected in both samples of a patient, whereas a private alteration was only found in one of the two samples. The two biopsies of the same tumour are labelled “TU1” and “TU2”. **(B)** Number of differentially expressed genes in the RNA-seq data set. **(C)** The most overrepresented Reactome pathways among the differentially expressed genes. The colour indicates the alteration status.

## 2.2. In-Depth Sequencing of 826 Selected Genes in 89 ccRCC Patients

In the second phase, the cohort was extended to 89 ccRCC patients. From each patient, two spatially separated biopsies of the primary tumour and a matched normal sample were collected. This data set includes 14 patients from the WES data set (two samples were removed due to an insufficient amount of material), and 75 additional patients. Utilising our customised gene panel, we sequenced the 178 tumour and 89 normal samples at high depth. The deep coverage enables the detection of low-frequency mutations, and the larger cohort provides increased statistical power such that rare subclonal mutation patterns can be detected. The sequenced reads contain unique molecular identifiers (UMIs), which allows for the correction of potential sequencing errors. The coverage of the panel sequencing (panel-seq) data set was on average 933x, and after UMI consensus building

and read filtering, it was 93x. The number of SNVs as well as indels in the panel-seq data set was on average 22 per patient (Figure 3, top panel). Pairwise comparison of the two biopsies from the same tumour revealed that on average, 62% of the mutations in a patient were private to one of the two samples with 87 of the 89 tumours (98%) containing at least one private mutation. Among the most frequently mutated genes, only 10 were mutated in more than 10% of the patients (only one in more than 50% of the patients), confirming the long-tail phenomenon commonly seen in cancer cohorts [16].

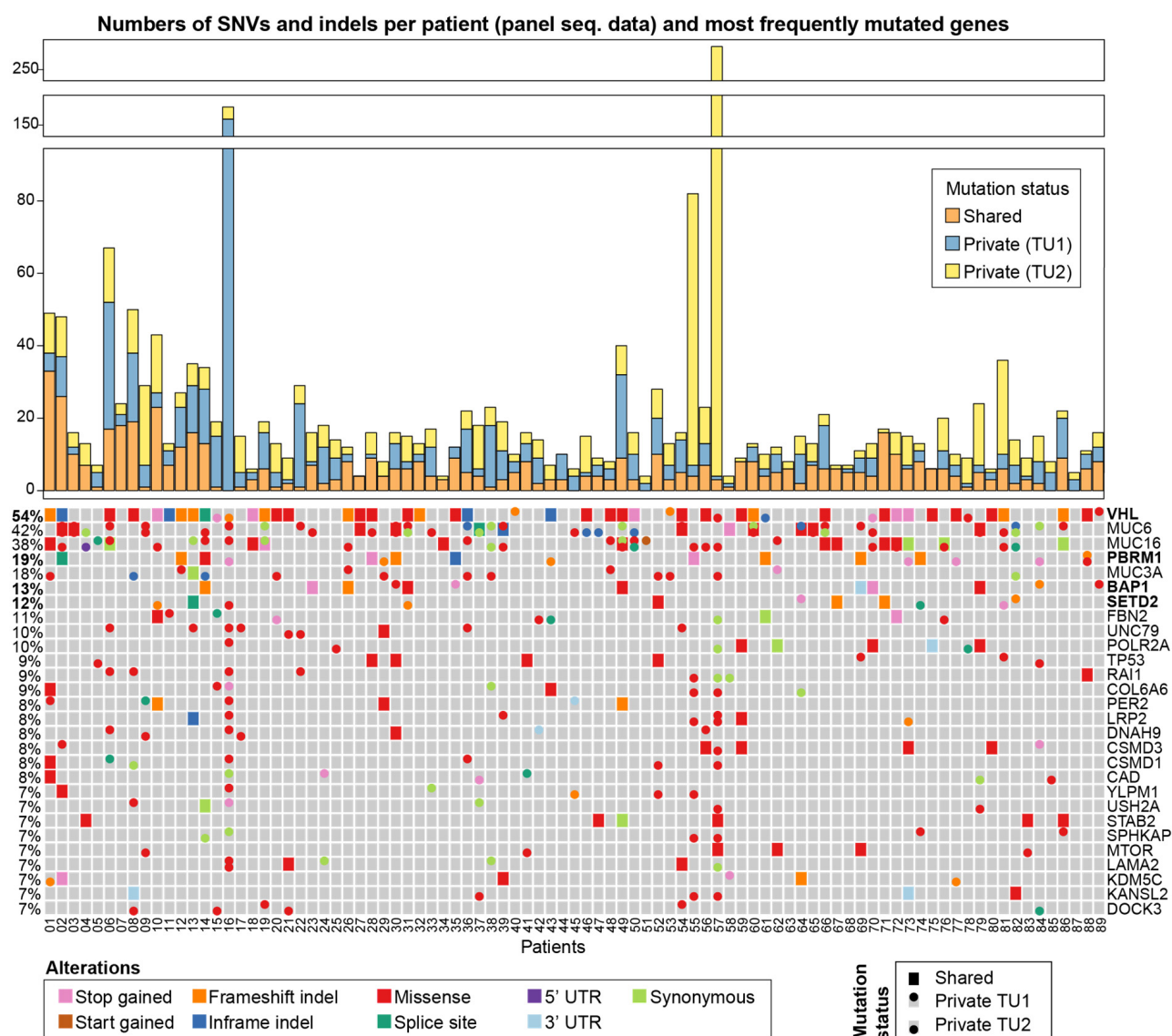
The four most commonly mutated genes in ccRCC, *VHL*, *PBRM1*, *SETD2* and *BAP1* [17], were also among the most frequently mutated genes in our data set (Figure 3, bottom panel). Three mucin genes: *MUC6* (42%), *MUC16* (38%), and *MUC3A* (18%) were also frequently mutated in our cohort. The mutation frequencies of *VHL*, *SETD2*, and *BAP1* were 54%, 12%, and 13%, respectively and comparable to the frequencies found in the TCGA cohort. Somewhat lower mutation frequencies were seen in *TP53* (9%), *mTOR* (7%), and *KDM5C* (7%), which were also reported in previous ccRCC studies [6,17,18]. Interestingly, in nine patients, the *VHL* mutations resided in only one of the two tumour samples. The mutations in *VHL* are known to occur early in tumour development [19,20], which is in line with our observation that in 39 of 48 (81%) cases, the *VHL* mutations were shared between both tumour biopsies of a patient. Private mutation was seen in all but two tumours; these mutations were observed in one of the two tumour samples underlining the strong genetic heterogeneity of ccRCC.

The number of mutations was unequally distributed between the two tumour biopsies in each case with identified private mutations; this is pronounced in patients 16, 55 and 57. Of the most frequently affected genes (Figure 3), three private mutations were found in *VHL*, *PBRM1*, and *SETD2* within TU2 of patient 16, whereas 14 private mutations were identified in TU1. A private mutation was also found in each tumour biopsy affecting the same gene: *LAMA2*. In patient 55, all seven private mutations in the most frequently mutated genes occurred in TU2, while in patient 57, 13 private mutations were found in TU2. Like patient 16, two private mutations in each tumour sample of patient 57 impacted the same gene; however, in patient 57, that gene is *LRP2*. Only five patients had no mutations in the most frequently affected genes ( $\geq 7\%$ ). In patient 88, the gene *PBRM1* was hit by different mutations in the two tumour samples, demonstrating a pattern of convergent phenotypic evolution where a gene is affected by multiple distinct mutations across the clones in the tumour. One tumour sample, TU1, had a frameshift deletion and a missense mutation, while the other tumour sample, TU2, had a missense mutation at a different locus in *PBRM1*. Both subclonal missense mutations of *PBRM1* are predicted to be deleterious according to the SIFT annotation [21]. Among the most subclonally affected genes in our data set were the mucins *MUC6*, *MUC16*, and *MUC3A*. A clonal exclusivity test was applied to the cohort of 89 ccRCC patients on the gene level. This test pinpoints the gene pairs that are mutated in the same patients but tend to be mutated in different subclones, hence are mutually exclusive on the level of subclones. The most striking gene pair was *TP53* and *MUC16*, which was clonally exclusive in patients 5 and 81 (Supplementary Figure S2, Supplementary Table S1).

### 2.3. Pathway-Level Clonal Exclusivity in 16 Ccrcc Patients

To reconstruct the evolutionary history of the tumours and assign mutations to specific clones, we used Cloe [23]. The WES data from paired tumour biopsies and matched normal samples of 16 ccRCC patients enabled us to map the mutated genes to pathways and to detect pathway pairs that are affected in several patients. The two most striking clonally exclusive pathway pairs, i.e., pathways that are aberrated in different clones of the same tumour in a mutually exclusive fashion, are “major pathway of rRNA processing in the nucleolus and cytosol” (referred to as pathway 1), “O-glycosylation of TSR domain-containing proteins” (pathway 2)}, and pathway 1, “defective B3GALT1 causes Peters-plus syndrome (PpS)” (pathway 3), which are clonally exclusive in both patients in which they are affected (Figure 4A,B, Supplementary Table S3), namely, patients 8 and 14 ( $p < 10^{-5}$ ).

Pathway 1 belongs to the category “metabolism of RNA”, while pathway 2 falls into the class “metabolism of proteins”, and pathway 3 is a disease pathway related to diseases of glycosylation [15]. Pathway 1 was also significantly enriched among the differentially expressed genes in 13 of 16 patients of this cohort.

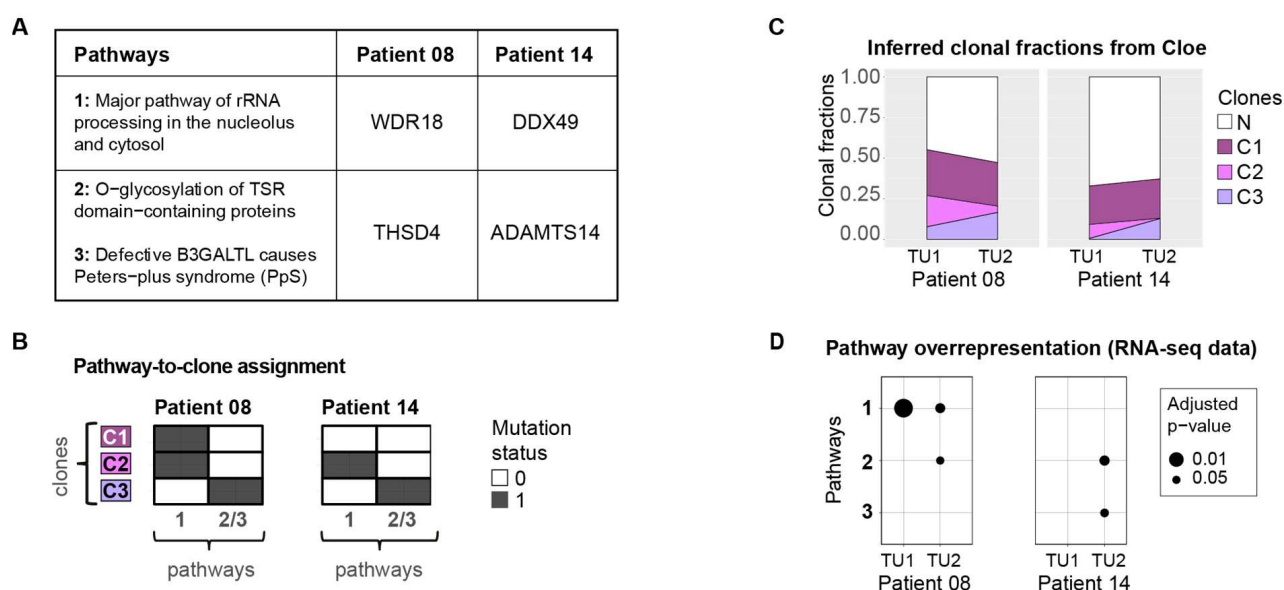


**Figure 3.** Genetic diversity in the panel seq data of 89 ccRCC patients. Top: numbers of shared and private mutations in the data set. Bottom: the heatmap highlights the mutations that were detected in the most frequently mutated genes. The four most frequently altered genes in ccRCC are *VHL*, *PBRM1*, *BAP1*, and *SETD2* [17] and are highlighted in bold. If a gene was hit by multiple mutations that all have the same status (shared, private TU1, or private TU2), the following ordering is applied to prioritise which colour is shown in the heatmap, starting with the highest priority: stop gained, start gained, frameshift indel, inframe indel, missense, splice site, five prime UTR, three prime UTR, synonymous. That means, if a gene has, e.g., a missense and a synonymous mutation, the missense mutation will be displayed in the heatmap. The variants were annotated with SnpEff [22] (Supplementary Table S2). Mutations in non-coding regions are omitted from the heatmap.

Aside from inferring the mutation-to-clone assignment, the Cloe software also estimates the fractions of the clones in each sample (Figure 4C). This is important in order to interpret possible changes on the transcriptomic level in the bulk samples. Pathway 1 was mapped to clones 1 and 2 in patient 8 (Figure 4B), which together have a clonal fraction of 47.5% and 30.7% in samples TU1, and TU2, respectively (Figure 4C). Pathway enrichment analysis shows that pathway 1 is also highly overrepresented in these two samples on



the transcriptomic level (Figure 4D). Pathways 2 and 3 were assigned to clone 3 which is, with 16.6%, the most abundant in sample TU2 of patient 8 (Figure 4C). Pathway 2 is also enriched among the differentially expressed genes in this sample, but pathway 3 is not suggesting that the underlying mutations seem to alter the expression of pathway 2 (Figure 4D). Pertaining to patient 14, no enrichment of pathway 1 could be found in either sample (Figure 4D). Pathway 1 was assigned to clone 2, which was estimated to have a clonal fraction of only 8.6% and 0.2% in samples TU1 and TU2, respectively (Figure 4C), which may explain why there was no signal detectable in the bulk transcriptome samples for this pathway. Pathways 2 and 3, however, were assigned to clone 3 (Figure 4B), which has an estimated clonal fraction of 0.5% and 12.8% in samples TU1 and TU2, respectively. In both bulk RNA samples, pathways 2 and 3 were enriched in the second tumour sample.



**Figure 4.** The two most striking clonally exclusive pathway pairs from the clonal exclusivity test when performed on the pathway level of the WES data set from 16 ccRCC patients. The pathway pairs are affected in patients 8 and 14 and in each patient, a different subset of genes is mutated. Hence, this clonal exclusivity pattern is only detectable on the pathway level. (A) The table displays the genes that are mutated in these pathways. (B) The heatmaps illustrate in which clones the genes in these pathways are mutated. (C) The proportion of cells from the clones in each of the two samples from the two patients. The label “N” represents the fraction of normal cells in the biopsy. (D) The data set also includes RNA-seq data from each sample. Among the differentially expressed genes in each sample, the pathways are significantly overrepresented in some of the samples.

### 3. Discussion

We analysed intra-tumour heterogeneity using two ccRCC patient cohorts. In the first phase, the investigation of the WES and RNA-seq data of 16 ccRCC patients revealed that intra-tumour heterogeneity is very pronounced on the genetic level, with an average of 40% of mutations found to be private. We also found that 31% of differentially expressed genes were detected in only one of the two patient tumour biopsies. In the second phase, the extended cohort of 178 tumour biopsies and the deep sequencing coverage enabled us to detect not only rare subclonal mutations, especially in *TP53* and mucin coding genes *MUC6*, *MUC16*, and *MUC3A*, but also early genomic alterations such as *PBRM1* and *VHL* with enough statistical power.

Intra-tumour heterogeneity has been reported previously by Gerlinger et al. in 2012. Following the extraction of 30 samples from four tumours, those authors observed up to 69% of somatic mutations not to be present within all samples [2]. In the same year, the analysis of 25 single cells from one ccRCC patient revealed the large extent of genetic heterogeneity between different tumour cells [24]. These studies of intra-tumour hetero-



geneity in ccRCC reported patterns of convergent phenotypic evolution in several genes including *VHL*, *BAP1*, *SETD2*, *PBRM1*, *PIK3CA*, *PTEN*, and *KDM5C* [2,4], which were also among the most frequently mutated genes in our cohort. ccRCC development is largely driven by the loss of one gene, *VHL*. Tumourigenesis typically starts with a large deletion on chromosome 3p, followed by mutational *VHL* inactivation. In addition to *VHL*, the 3p deletion also removes one copy of *PBRM1*, *BAP1*, and *SETD2*. Since *VHL* inactivation alone is insufficient [25], mutations in *PBRM1* and *BAP1* are necessary for ccRCC development. Importantly, these mutations tend to be mutually exclusive [19]. Interestingly, we have seen *VHL* mutations in only one of the two tumour samples in nine out of eighty-nine patients. This implies that these alterations would have been missed if the other tumour piece was analysed alone, with the consequence that these tumours would have been considered as *VHL* wild-type ccRCC. While *VHL* mutations are considered truncal, we hypothesize that in these nine samples, the mutated allele was lost in the respective subclone [26–28].

In addition, some of the genes were affected by multiple distinct mutations across the clones in the tumour. This is comparable to some of our previous findings in the *VHL* gene [3]. *PBRM1* mutations occur in 19% of the patients in our cohort, which is less than the frequency reported from TCGA [17,18]. We found a pattern of convergent phenotypic evolution in *PBRM1*: the gene was hit by two different deleterious missense mutations in each of the two biopsies of one patient.

We identified frequent subclonal mutations in *MUC6*, *MUC16*, and *MUC3A*, indicating that alterations of these mucin genes may also be critical in ccRCC development. In the TCGA data set, the mucins *MUC4*, and *MUC16* were also among the seven most frequently mutated genes [17], but our mutation frequencies of *MUC6*, *MUC16*, and *MUC3A* are 42%, 38%, and 18%, higher than those reported in TCGA [17]. In previous ccRCC studies, *MUC16* was reported to be among the most recurrently mutated genes [18,29]. Our analysis of the mutation distribution within the *MUC16* (also termed CA125) amino acid sequence revealed that non-silent mutations seem to be clustering at the end of the sequence in the SEA domains (Supplementary Figure S3) whose precise function is not known. These extracellular SEA domains can be extensively O-glycosylated and it was suggested that they can bind nucleic acids or sugars, or be released through cleavage [30]. In this context, it is of note that several RCC studies reported an association between increased levels of *MUC16*, poor prognosis and advanced tumour stage, suggesting the potential use of *MUC16* as a serum biomarker in RCC [31–33]. Although *MUC16* had one of the highest mutation frequencies, our immunohistochemistry analysis showed no expression of *MUC16* in the ccRCC samples (data not shown, see methods for IHC protocol). How mutations in SEA domains affect *MUC16*'s function and contribute to its release in the serum in ccRCC patients remains to be evaluated. It was shown that in lung cancer, *MUC16* mutations can lead to its oncogenic upregulation [34] and the overexpression of *MUC16* is associated with increased tumour cell growth, cancer cell migration, and resistance to cytotoxic drugs [35]. Recent studies also discovered frequent non-silent *MUC16* mutations in breast cancer [36,37], another cancer type in which *MUC16* was observed to be overexpressed [38]. Furthermore, *MUC16* mutations have been implicated as cancer-driving in a pan-cancer analysis that assessed the functional impact of mutations on differential gene expression profiles [39].

The majority of mucin gene mutations found in our cohort have a low variant allele frequency (VAF). Specifically, more than 75% of the mutations in *MUC6*, *MUC16*, and *MUC3A* have a VAF below 10%, and almost half below 5%. Given the mostly low VAFs of mutated mucin genes in our study, the analysis of the effect of mutations on mucin protein expression and its prognostic value would be very challenging. *MUC6* and *MUC16* protein expression is hardly detectable in ccRCC and *MUC3A* show weak to moderate expression in all ccRCC analysed (Human Protein Atlas and own data (*MUC16*), not shown). Despite the difficult interpretation of varying positivity of *MUC3A*, an increased expression of this mucin was correlated with poor prognosis in localised ccRCC [40].

Besides *MUC6*, *MUC16* and *MUC3A*, additional mucins may play an important role in ccRCC, as their expression level was shown to be predictive of clinical outcome. Decreased expression of Mucin 4 and Mucin 18 predicted poor prognosis [41,42] whereas high Mucin 7, Mucin 5A, and Mucin 13 expression was found to be associated with worse patient outcome [43–45].

*TP53*, a well-known tumour suppressor, was found to be mutated in less than 10% of ccRCC [17,18,46,47], which is confirmed in the cohort analysed here (9%). *TP53* mutations were associated with the reduced survival of renal cancer patients [48,49]. Of note, a previous study of the intra-tumour diversity in ten ccRCC cases revealed that mutation in *TP53* were one of the most extreme examples of gene mutations being detected more often when sequencing multiple biopsies per tumour instead of a single one [4]. We confirmed this finding as four of eight mutations were detected in only one of the two tumour samples. This observation suggests that *TP53* mutations may be a crucial subclonal event in ccRCC and explains the low prevalence of *TP53* mutations in earlier studies. Motzer et al. have evaluated somatic alterations across different histological subtypes and reported a lower prevalence of *PBRM1* mutations in ccRCC with sarcomatoid differentiation, whereas *TP53* mutations had an increased prevalence in non-ccRCC with sarcomatoid differentiation. Sarcomatoid RCC exhibited a highly proliferative phenotype with high immune presence and PD-L1 expression, explaining increased sensitivity to therapeutic intervention with atezolizumab+bevacizumab versus sunitinib [50,51]. Similar to non-ccRCC with sarcomatoid differentiation, subclonal *TP53* mutations could also be a first molecular step into the development of an aggressive phenotype of ccRCC leading to sarcomatoid differentiation. Motzer et al. have recently identified seven ccRCC subtypes with specific angiogenesis, immune, metabolic, stromal, and cell-cycle profiles showing differential clinical outcomes to VEGF blockade alone or in combination with anti-PD-L1 [14]. These molecular clades have a differential prevalence of *TP53*, *PBRM1*, *KDM5C*, and *CDKN2A/2B* alterations. Our observation of subclonal *TP53* mutations suggests that primarily tissue samples with sarcomatoid differentiation may display high levels of intra-tumour heterogeneity.

A total of 87 tumours (98%) had private mutations and were detectable in only one of two tumour samples. Given the relatively large tumour volumes of ccRCC, with pT1 and pT2 tumours having diameters of up to 7 and 11 cm, respectively, spatial heterogeneity represents a tremendous challenge to individual therapy. Underrating the mutational burden due to spatial heterogeneity of gene alterations is thus a common problem in cancer research as well as in molecular tumour diagnostics. Single cell analysis and appropriate bioinformatics tools may help to overcome this bottleneck, particularly if only single little tumour biopsies are available.

Within a subset of patients, we see a large difference in mutational counts between the two biopsies. These differences could arise due to differences in tumour purity or perhaps aberrations affecting genes needed for DNA damage repair, resulting in an accumulation of mutations [52]. We controlled for tumour purity with our requirement of at least 70% tumour cells and so explored the potential aberrations in genes involved in DNA repair obtained from MutSigDB [53]. Missense mutations were found in both *ERCC1* and *POLR2A* in TU1 of patient 16. *ERCC1*, together with XPF, forms a nuclease essential for nucleotide excision repair with *ERCC1* required for DNA binding [54,55]. *POLR2A* encodes the largest subunit of RNA polymerase II (RNAPII). RNAPII initiates the recruitment of transcription-coupled nucleotide excision repair factors such as CSB when stalling at DNA lesions blocking translation [56]. Dysfunction in just one of these proteins may lead to an increase in the number of mutations, as seen in TU1 of patient 16. In TU2 of patient 55, we identified downstream intron variants of *XPC* that produce a protein of the same name functioning to recognise DNA damage during the global genome–nucleotide excision repair pathway [57]. Although these *XPC* variants have not been classed as altering *XPC* function, given the large difference seen between TU1 and TU2 in patient 55, these variants may be a cause of the accumulation of mutations observed in the second biopsy.

In an attempt to identify co-existing clones with affected pathways related to the metabolism of proteins and RNA, we applied a clonal exclusivity test to 16 ccRCC patients on the pathway level. This test allows the identification of pathways that are perturbed in different clones of the same tumour in a mutually exclusive fashion. These pathway alterations occurred in clones that evolved in parallel along different branches of the tumour phylogeny. The two most striking pathway pairs include “major pathway of rRNA processing in the nucleolus and cytosol” (pathway 1), which is clonally exclusive with “O-glycosylation of TSR domain-containing proteins” (pathway 2) and “defective B3GALTL causes Peters-plus syndrome (PpS)” (pathway 3) (Figure 4A). The pathway pairs {1, 2} and {1, 3} are affected in the two patients through a different subset of genes (Figure 4A), namely *WDR18* and *THSD4* in patient 8, and *DDX49* and *ADAMTS14* in patient 14. *ADAMTS14* belongs to the ADAMTS protein family, which are secreted zinc metalloproteases that play a role in the extracellular matrix related to angiogenesis and cancer [58]. *THSD4* is also referred to as ADAMTS-like protein 6 and is also secreted to the extracellular matrix [59]. Both *ADAMTS14* and *THSD4* contain the thrombospondin type 1 repeat (TSR) domain [60]. The proteins with TSR domains can undergo O-fucosylation, a protein modification that plays a role in angiogenesis and Notch signalling [60–62]. To conclude, the pathways detected as clonally exclusive (pathways 1, 2, and 3) are also enriched among the differentially expressed genes in some of the samples. Pathways 2 and 3 are functionally deregulated in only one of the two biopsies, showing that this deregulation is subclonal. The deregulation may arise due to the mutations in *THSD4* and *ADAMTS14*, which are members of these pathways. Whilst being of interest, the exclusivity pattern was observed in two patients, and therefore validation of these findings in a larger cohort would be beneficial.

## 4. Materials and Methods

### 4.1. Experimental Design

The analysis of intra-tumour heterogeneity and subclonal mutation patterns was comprised of two phases (Figure 1). In an initial discovery phase, whole-exome and transcriptome sequencing data from paired tumour biopsies from 16 ccRCC patients plus one matched normal sample per patient were analysed to obtain an overview of the diversity in these samples. In this first exploratory step, the detected mutations informed the design of our gene panel for the second phase. Furthermore, frequently mutated ccRCC genes from the publicly available data sets provided by the Cancer Genome Atlas Research Network (TCGA) [17] were considered for the selection of genes in the panel. During the second phase, a total of 826 selected genes were then targeted at deep coverage in an extended cohort of 89 patients for a detailed analysis of tumour heterogeneity.

### 4.2. Patient Material

Two cohorts of 16 and 89 ccRCC patients with no prior treatment were chosen for the sequence analyses. The tumours of these ccRCC patients were classified according to the 2016 WHO classification [63] and reviewed by H.M. From each patient, two tumour samples and one matched normal tissue were selected. From each frozen and FFPE tissue block, haematoxylin and eosin stained sections were prepared and reviewed by a pathologist (H.M.) to ensure tissue integrity. Only tumours with at least 70% tumour cells were included in our cohort. For whole exome and RNA sequencing of 48 (16 × 3) tissue samples, 5–10 frozen sections (10 µm) were used for DNA and RNA extraction. For in depth sequencing, 3 punches (0.6 mm diameter) were taken from 267 (89 × 3) formalin-fixed, paraffin-embedded tissue blocks. All tissue samples were anonymised.

### 4.3. Whole Exome Sequencing

The first data set encompasses two spatially separated primary tumour biopsies and one matched normal sample from each of the sixteen clear cell renal cell carcinoma (ccRCC) patients. The whole exome was sequenced using the Illumina HiSeq 2000 system to obtain

101-bp paired-end reads. The computational pipeline to analyse the data was a customised version of the NGS-pipe framework [64] that included the following steps: adapter clipping and trimming of low-quality bases with Trimmomatic [65], alignment of the reads to the human reference genome version hg19 using bwa [66], and read processing with SAMtools [67], Picard tools [68], and bam-readcount. Reads were realigned locally around indels, and base qualities were recalibrated with the Genome Analysis Toolkit (GATK) [69]. Single-nucleotide variants (SNVs) were called using the rank-combination [70] of deepSNV [3], JointSNVMix2 [71], MuTect [72], SiNVICT [73], Strelka [74], and VarScan2 [75]. The rank-combination is a method that combines the results of different variant callers by integrating the ranked lists of variants to generate a combined ranking [70]. P-values of deepSNV were corrected for multiple testing with the R package IHW [76]. Indels were called using SiNVICT [73], Strelka [74], VarDict [77], and VarScan2 [75], and combining them with the rank-combination [70]. For copy number variant detection, the tool Sequenza [78] was employed. Mutations in copy number neutral regions were selected as input for Cloe [23] in order to reconstruct the evolutionary history of a tumour and to assign the mutations to different clones. In order to account for the uncertainty in the phylogenetic tree inference, Cloe was run 20 times with different seeds.

#### 4.4. Transcriptomic Data Generation and Analysis

Paired-end and single-end RNA-sequencing was performed on the Illumina HiSeq 2000 system to generate 101-bp paired-end reads, and 51-bp single-end reads for the 48 samples from the initial 16 patients. For the computational analysis, the NGS-pipe framework was adapted [64]. Reads were clipped and trimmed using Trimmomatic [65], and alignment was performed with STAR [79]. Read counts for the genes were obtained with the program featureCounts [80]. Differential gene expression analysis was carried out using DESeq2 [81], comparing each tumour sample to its paired normal sample using both single and paired-end data (i.e., a 2 vs. 2 design). Genes with a q-value less than 0.01 were considered differentially expressed. The R package WebGestaltR [82] was applied to perform enrichment analysis using all differentially expressed genes (up- and downregulated) together. As a background gene list for the enrichment analysis, only expressed genes were included. More precisely, in each comparison of a patient's tumour samples to the matched normal samples, the expressed genes were included in the background gene list if they had at least a count of 10 fragments across the tumour and normal samples.

#### 4.5. Panel Sequencing

The second data set is a panel sequencing data set. It comprises an extended cohort of patients from which a selected set of genes were sequenced at higher depth. The selection of 826 genes was informed by the mutated genes detected in the WES data set, as well as from the frequently mutated ccRCC genes in TCGA [17] (see Additional Files S3 and S4 for the 826 gene list and bed file). We generated panel-seq data from 89 ccRCC patients, including two spatially separated primary tumour biopsies, and a matched normal sample per patient. This data set comprises 14 of the patients from the WES data set, and 75 additional patients. The data were sequenced using the Illumina HiSeq 2500 system. The sequenced reads contain unique molecular identifiers (UMIs), and this allows for the correction of potential sequencing errors. Reads with identical UMIs, which are mapped to the same genomic position, come from the same DNA molecule, and therefore, the consensus sequence can be built, and the variants can be called with higher confidence.

The pipeline for analysing the sequencing data was again a customised version of the NGS-pipe framework [64] including the following steps: raw reads were clipped and trimmed using the tool SeqPurge [83]. Reads were aligned to the human reference genome version hg19 with the aligner bwa [66]. Reads were further processed using SAMtools [67], Picard tools [68] and bam-readcount. Local realignment around indels was carried out with GATK [69]. We used the software UMI-tools [84] to group reads with identical UMI and identical mapping position together, and an in-house tool to build the consensus sequence



and thereby correct sequencing errors. Our in-house tool takes the grouped reads with identical UMIs and identical mapping positions and attempts to generate the consensus sequence from these grouped reads. If the reads contain contradicting bases at a nucleotide position, it is masked with the base “N”. The SNV and indel calling was similar as for the WES data set. Some of our samples are from formalin-fixed paraffin-embedded (FFPE) material. FFPE samples are known to harbour artificial C > T and G > A alterations [85,86]. They occur mostly at lower frequencies in the range between 1–10% variant allele frequency (VAF), since the DNA damage occurs at different genomic positions in different cells [85,87]. To remove potential FFPE artefacts, we filtered out C > T and G > A mutations that had a VAF < 10%. The tool Cloe [23] used for the tree inference requires as input mutations in copy number-neutral regions. In order to filter out mutations that are in potential copy number variant regions, mutations that are within 4000 bps of an imbalanced heterozygous germline SNP were filtered out. An imbalanced heterozygous germline SNP is a SNP that has a VAF between 40–60% in the normal sample, but in the tumour sample the VAF is out of these bounds, indicating a potential copy number change. Finally, in order to perform quality control we used Qualimap [88] and FastQC [89] in the WES and panel sequencing data sets.

#### 4.6. Testing for Pathway-Level Clonal Exclusivity

The mutations detected in the WES data set were assigned to clones with Cloe [23], mapped to genes, and subsequently, the genes were mapped to pathways using the Reactome pathway database [15]. This procedure resulted in a total of 877 affected pathways.

For functional annotation of the variants, SnpSift [90] and SnpEff [22], as well as the data bases COSMIC [47] version 80, and dbSNP [91] version 138 were used. In order to identify pathways that are altered in a clonally exclusive fashion, we employed the statistical test implemented in GeneAccord [92]. We only kept genes with a potential impact for this analysis. While mutations such as synonymous or intronic variants are informative for the tree inference, they were not of interest for the clonal exclusivity test. Non-silent mutations are more likely to change the phenotype of the clones and therefore these mutations are potential candidates for inducing clonal interactions. For the estimation of background rates of clonal exclusivity and co-occurrence, it is therefore important to focus on non-silent mutations in order to have accurate estimates of their clonally exclusive background distribution. To filter out silent mutations, we used the annotation program SnpEff, which classifies variants into four categories based on the potential impact of the mutation [22]. These are, in descending order of importance, the categories “HIGH”, “MODERATE”, “LOW”, and “MODIFIER”. Examples of the category “HIGH” would be frameshift indels. Missense mutations and inframe indels are classified as “MODERATE”. The category “LOW” includes synonymous and splice region mutations. Variants that are annotated as “upstream”, “intronic”, or “UTR region” fall into the class “MODIFIER”. For the GeneAccord-based clonal exclusivity analysis, we kept mutations that are in the category “HIGH”, “MODERATE”, and from the class “LOW” we kept all variants with the exception of: synonymous variants, or mutations that are annotated as the case where a start codon mutates into another start codon, or analogous for stop codon. To sum up, we kept variants such as missense, frameshift or inframe indel or variants in splice regions, but filtered out variants that are synonymous, intronic or in the UTR regions.

#### 4.7. Statistical Analysis

For the data analysis in R [93] as well as for visualising results, several R packages were used including biomaRt [94,95], caTools [96], dplyr [97], ggplot2 [98], ggpubr [99], gtools [100], maxLik [101], tibble [102], magrittr [103], reshape2 [104], RColorBrewer [105], ComplexHeatmap [106], and survival [107].

#### 4.8. Immunohistochemistry

TMA sections (2.5 µm) on glass slides were subjected to immunohistochemical analysis stained using Ultra Discovery (Ventana, Roche Diagnostics, Rotkreuz, Switzerland). MUC16/CA125 was immunostained using monoclonal mouse anti-MUC16 antibody (clone X75, cat. no. M1-90039; Invitrogen, diluted 1:1000 in Bond medium). MUC16 was made visible using IHC Refine kits (Ventana). Normal and tumour tissue (cut off: >5% tumour cells) were considered MUC16-positive if tumour cells showed unequivocal weak, moderate or strong cytoplasmic and membranous expression.

#### 5. Conclusions

In summary, the systematic analysis of the clone constellations as performed here in large patient cohorts will contribute towards a better understanding of the evolutionary forces beyond mutation and selection that drive tumour evolution and will help to improve treatment strategies available for those with ccRCC.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13092163/s1>, Figure S1: Euclidean distances between samples of each patient, Figure S2: Number of mutated genes per patient used as input for GeneAccord & average rates of clonal exclusivity per patient, Figure S3: MUC16 mutations, Table S1: The 16 gene pairs with a positive parameter delta of the clonal exclusivity test applied to the 89 ccRCC, Table S2: The variant categories from Figure 3 and the corresponding annotations from SnpEff, Table S3: The ten most striking pathway pairs with a positive parameter delta of the clonal exclusivity test applied to the 16 ccRCC on the pathway level, Additional File 2: Euclidean distance matrices (.txt), Additional File 3: 826 gene panel, Additional File 4: BED file corresponding to gene panel.

**Author Contributions:** N.B., C.B., P.S., and H.M. formulated the research goals and administrated the project. N.B., C.B., and H.M. acquired the funding. P.S. and H.M. collected the ccRCC samples. C.B. and E.B. performed the sequencing. J.S. and A.L.M. implemented the computational pipeline for the genomic sequencing data analysis. A.L.M. performed the computational analysis of the genomic and transcriptomic sequencing data. A.L.M. conducted the statistical analyses for differential gene expression and clonal exclusivity. J.K., N.B., C.B., P.S., H.M. provided supervision. A.L.M. visualised the results. A.L.M. and A.A.B. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** Part of this work was supported by SystemsX.ch IPhD Grant SXPHI0\_142005, ERC Synergy Grant 609883, Horizon 2020 SOUND Grant Agreement no 633974, and SNF grant 310030\_166391 (to HM).

**Institutional Review Board Statement:** This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local commission of ethics (BASEC\_2019-01959). The retrospective use of sequencing data obtained from normal and tumour tissues of ccRCC patients is in accordance with the Swiss Law ("Humanforschungsgesetz").

**Informed Consent Statement:** Patient consent was waived as according to Article 34 of the Humanforschungsgesetz, the use of biomaterial and patient data for research purposes is permitted without informed consent under certain conditions that include the present cases. Law abidance of this study was reviewed and approved by the ethics commission of the Canton Zurich.

**Data Availability Statement:** The generated read data from the clear cell renal cell carcinoma samples and the matched normal samples have been deposited in the European Nucleotide Archive under accession numbers ERP108328 and ERP108326. Full pathway-by-clone matrices from each patient have been uploaded to the Github repository ([https://github.com/cbg-ethz/GeneAccord/tree/master/data/clone\\_pws\\_tbl\\_16\\_WES](https://github.com/cbg-ethz/GeneAccord/tree/master/data/clone_pws_tbl_16_WES)).

**Acknowledgments:** The authors would like to thank Francesco Marass for support with visualising the variant allele frequencies and Susanne Dettwiler from the Tissue Biobank USZ for providing tissue material. The graphical abstract was created with BioRender.com.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **1976**, *194*, 23–28. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gerlinger, M.; Rowan, A.J.; Horswell, S.; Math, M.; Larkin, J.; Endesfelder, D.; Gronroos, E.; Martinez, P.; Matthews, N.; Stewart, A.; et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **2012**, *366*, 883–892. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gerstung, M.; Beisel, C.; Rechsteiner, M.; Wild, P.; Schraml, P.; Moch, H.; Beerenwinkel, N. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat. Commun.* **2012**, *3*, 811. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gerlinger, M.; Horswell, S.; Larkin, J.; Rowan, A.J.; Salm, M.P.; Varela, I.; Fisher, R.; McGranahan, N.; Matthews, N.; Santos, C.R.; et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **2014**, *46*, 225–233. [\[CrossRef\]](#) [\[PubMed\]](#)
- Beerenwinkel, N.; Greenman, C.D.; Lagergren, J. Computational cancer biology: An evolutionary perspective. *PLoS Comput. Biol.* **2016**, *12*, e1004717. [\[CrossRef\]](#)
- Turajlic, S.; Xu, H.; Litchfield, K.; Rowan, A.; Horswell, S.; Chambers, T.; O'Brien, T.; Lopez, J.I.; Watkins, T.B.K.; Nicol, D.; et al. Deterministic evolutionary trajectories influence primary tumor growth: Tracerx renal. *Cell* **2018**, *173*, 595–610. [\[CrossRef\]](#)
- Moch, H.; Schraml, P.; Bubendorf, L.; Richter, J.; Gasser, T.C.; Mihatsch, M.J.; Sauter, G. Intratumoral heterogeneity of von Hippel-Lindau gene deletions in renal cell carcinoma detected by fluorescence in situ hybridization. *Cancer Res.* **1998**, *58*, 2304–2309.
- Martinez, P.; Birkbak, N.J.; Gerlinger, M.; McGranahan, N.; Burrell, R.A.; Rowan, A.J.; Joshi, T.; Fisher, R.; Larkin, J.; Szallasi, Z.; et al. Parallel evolution of tumour subclones mimics diversity between tumours. *J. Pathol.* **2013**, *230*, 356–364. [\[CrossRef\]](#)
- Turajlic, S.; Xu, H.; Litchfield, K.; Rowan, A.; Chambers, T.; Lopez, J.I.; Nicol, D.; O'Brien, T.; Larkin, J.; Horswell, S.; et al. Tracking cancer evolution reveals constrained routes to metastases: Tracerx renal. *Cell* **2018**, *173*, 581–594. [\[CrossRef\]](#)
- Okegawa, T.; Morimoto, M.; Nishizawa, S.; Kitazawa, S.; Honda, K.; Araki, H.; Tamura, T.; Ando, A.; Satomi, Y.; Nutahara, K.; et al. Intratumor Heterogeneity in Primary Kidney Cancer Revealed by Metabolic Profiling of Multiple Spatially Separated Samples within Tumors. *EBioMedicine* **2017**, *19*, 31–38. [\[CrossRef\]](#)
- Choueiri, T.K.; Motzer, R.J. Systemic Therapy for Metastatic Renal-Cell Carcinoma. *N. Engl. J. Med.* **2017**, *376*, 354–366. [\[CrossRef\]](#)
- Penticuff, J.C.; Kyprianou, N. Therapeutic challenges in renal cell carcinoma. *Am. J. Clin. Exp. Urol.* **2015**, *3*, 77–90. [\[PubMed\]](#)
- Deleuze, A.; Saout, J.; Dugay, F.; Peyronnet, B.; Mathieu, R.; Verhoest, G.; Bensalah, K.; Crouzet, L.; Laguerre, B.; Belaud-Rotureau, M.-A.; et al. Immunotherapy in renal cell carcinoma: The future is now. *Int. J. Mol. Sci.* **2020**, *21*, 2532. [\[CrossRef\]](#) [\[PubMed\]](#)
- Motzer, R.J.; Banchereau, R.; Hamidi, H.; Powles, T.; McDermott, D.; Atkins, M.B.; Escudier, B.; Liu, L.-F.; Leng, N.; Abbas, A.R.; et al. Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. *Cancer Cell* **2020**, *38*, 803–817. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2018**, *46*, D649–D655. [\[CrossRef\]](#)
- Garraway, L.A.; Lander, E.S. Lessons from the cancer genome. *Cell* **2013**, *153*, 17–37. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **2013**, *499*, 43–49. [\[CrossRef\]](#)
- Sato, Y.; Yoshizato, T.; Shiraishi, Y.; Maekawa, S.; Okuno, Y.; Kamura, T.; Shimamura, T.; Sato-Otsubo, A.; Nagae, G.; Suzuki, H.; et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **2013**, *45*, 860–867. [\[CrossRef\]](#) [\[PubMed\]](#)
- Peña-Llopis, S.; Christie, A.; Xie, X.-J.; Brugarolas, J. Cooperation and antagonism among cancer genes: The renal cancer paradigm. *Cancer Res.* **2013**, *73*, 4173–4179. [\[CrossRef\]](#)
- Mitchell, T.J.; Turajlic, S.; Rowan, A.; Nicol, D.; Farmery, J.H.R.; O'Brien, T.; Martincorena, I.; Tarpey, P.; Angelopoulos, N.; Yates, L.R.; et al. Timing the landmark events in the evolution of clear cell renal cell cancer: Tracerx renal. *Cell* **2018**, *173*, 611–623. [\[CrossRef\]](#)
- Sim, N.-L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [\[CrossRef\]](#)
- Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [\[CrossRef\]](#) [\[PubMed\]](#)
- Marass, F.; Mouliere, F.; Yuan, K.; Rosenfeld, N.; Markowitz, F. A phylogenetic latent feature model for clonal deconvolution. *Ann. Appl. Stat.* **2016**, *10*, 2377–2404. [\[CrossRef\]](#)
- Xu, X.; Hou, Y.; Yin, X.; Bao, L.; Tang, A.; Song, L.; Li, F.; Tsang, S.; Wu, K.; Wu, H.; et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **2012**, *148*, 886–895. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gu, Y.-F.; Cohn, S.; Christie, A.; McKenzie, T.; Wolff, N.; Do, Q.N.; Madhuranthakam, A.J.; Pedrosa, I.; Wang, T.; Dey, A.; et al. Modeling renal cell carcinoma in mice: Bap1 and pbrm1 inactivation drive tumor grade. *Cancer Discov.* **2017**, *7*, 900–917. [\[CrossRef\]](#)
- Bissig, H.; Richter, J.; Desper, R.; Meier, V.; Schraml, P.; Schäffer, A.A.; Sauter, G.; Mihatsch, M.J.; Moch, H. Evaluation of the clonal relationship between primary and metastatic renal cell carcinoma by comparative genomic hybridization. *Am. J. Pathol.* **1999**, *155*, 267–274. [\[CrossRef\]](#)

27. Dagher, J.; Kammerer-Jacquet, S.-F.; Brunot, A.; Pladys, A.; Patard, J.-J.; Bensalah, K.; Perrin, C.; Verhoest, G.; Mosser, J.; Lespagnol, A.; et al. Wild-type VHL Clear Cell Renal Cell Carcinomas Are a Distinct Clinical and Histologic Entity: A 10-Year Follow-up. *Eur. Urol. Focus* **2016**, *1*, 284–290. [\[CrossRef\]](#)
28. Batavia, A.A.; Schraml, P.; Moch, H. Clear cell renal cell carcinoma with wild-type von Hippel-Lindau gene: A non-existent or new tumour entity? *Histopathology* **2019**, *74*, 60–67. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Arai, E.; Sakamoto, H.; Ichikawa, H.; Totsuka, H.; Chiku, S.; Gotoh, M.; Mori, T.; Nakatani, T.; Ohnami, S.; Nakagawa, T.; et al. Multilayer-omics analysis of renal cell carcinoma, including the whole exome, methylome and transcriptome. *Int. J. Cancer* **2014**, *135*, 1330–1342. [\[CrossRef\]](#)
30. Maeda, T.; Inoue, M.; Koshihara, S.; Yabuki, T.; Aoki, M.; Nunokawa, E.; Seki, E.; Matsuda, T.; Motoda, Y.; Kobayashi, A.; et al. Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16). *J. Biol. Chem.* **2004**, *279*, 13174–13182. [\[CrossRef\]](#)
31. Grankvist, K.; Ljungberg, B.; Rasmuson, T. Evaluation of five glycoprotein tumour markers (CEA, CA-50, CA-19-9, CA-125, CA-15-3) for the prognosis of renal-cell carcinoma. *Int. J. Cancer* **1997**, *74*, 233–236. [\[CrossRef\]](#)
32. Lucarelli, G.; Ditunno, P.; Bettocchi, C.; Vavallo, A.; Rutigliano, M.; Galleggiante, V.; Larocca, A.M.V.; Castellano, G.; Gesualdo, L.; Grandaliano, G.; et al. Diagnostic and prognostic role of preoperative circulating CA 15-3, CA 125, and beta-2 microglobulin in renal cell carcinoma. *Dis. Markers* **2014**, *2014*, 689795. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Bamias, A.; Chorti, M.; Deliveliotis, C.; Trakas, N.; Skolarikos, A.; Protogerou, B.; Legaki, S.; Tsakalou, G.; Tamvakis, N.; Dimopoulos, M.A. Prognostic significance of CA 125, CD44, and epithelial membrane antigen in renal cell carcinoma. *Urology* **2003**, *62*, 368–373. [\[CrossRef\]](#)
34. Kanwal, M.; Ding, X.-J.; Song, X.; Zhou, G.-B.; Cao, Y. MUC16 overexpression induced by gene mutations promotes lung cancer cell growth and invasion. *Oncotarget* **2018**, *9*, 12226–12239. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Lakshmanan, I.; Salfity, S.; Seshacharyulu, P.; Rachagani, S.; Thomas, A.; Das, S.; Majhi, P.D.; Nimmakayala, R.K.; Vengoji, R.; Lele, S.M.; et al. MUC16 Regulates TSPYL5 for Lung Cancer Cell Growth and Chemoresistance by Suppressing p53. *Clin. Cancer Res.* **2017**, *23*, 3906–3917. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Bareche, Y.; Venet, D.; Ignatiadis, M.; Aftimos, P.; Piccart, M.; Rothe, F.; Sotiriou, C. Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann. Oncol.* **2018**, *29*, 895–902. [\[CrossRef\]](#)
37. Sachs, N.; de Ligt, J.; Kopper, O.; Gogola, E.; Bounova, G.; Weeber, F.; Balgobind, A.V.; Wind, K.; Gracanin, A.; Begthel, H.; et al. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell* **2018**, *172*, 373–386. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Lakshmanan, I.; Ponnusamy, M.P.; Das, S.; Chakraborty, S.; Haridas, D.; Mukhopadhyay, P.; Lele, S.M.; Batra, S.K. MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. *Oncogene* **2012**, *31*, 805–817. [\[CrossRef\]](#)
39. Cai, C.; Cooper, G.; Lu, K.; Ma, X.; Xu, S.; Zhao, Z.; Chen, X.; Xue, Y.; Lee, A.; Clark, N.; et al. Systematic Discovery of the Functional Impact of Somatic Genome Alterations in Individual Tumors through Tumor-specific Causal Inference. *BioRxiv* **2018**. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Niu, T.; Liu, Y.; Zhang, Y.; Fu, Q.; Liu, Z.; Wang, Z.; Fu, H.; Xu, J.; Liu, K. Increased expression of MUC3A is associated with poor prognosis in localized clear-cell renal cell carcinoma. *Oncotarget* **2016**, *7*, 50017–50026. [\[CrossRef\]](#)
41. Bai, Q.; Liu, L.; Long, Q.; Xia, Y.; Wang, J.; Xu, J.; Guo, J. Decreased expression of mucin 18 is associated with unfavorable postoperative prognosis in patients with clear cell renal cell carcinoma. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 11005–11014. [\[PubMed\]](#)
42. Fu, H.; Liu, Y.; Xu, L.; Chang, Y.; Zhou, L.; Zhang, W.; Yang, Y.; Xu, J. Low Expression of Mucin-4 Predicts Poor Prognosis in Patients with Clear-Cell Renal Cell Carcinoma. *Medicine* **2016**, *95*, e3225. [\[CrossRef\]](#) [\[PubMed\]](#)
43. NguyenHoang, S.; Liu, Y.; Xu, L.; Chang, Y.; Zhou, L.; Liu, Z.; Lin, Z.; Xu, J. High mucin-7 expression is an independent predictor of adverse clinical outcomes in patients with clear-cell renal cell carcinoma. *Tumour Biol.* **2016**, *37*, 15193–15201. [\[CrossRef\]](#)
44. Zhang, H.; Liu, Y.; Xie, H.; Liu, W.; Fu, Q.; Yao, D.; Xu, J.; Gu, J. High mucin 5AC expression predicts adverse postoperative recurrence and survival of patients with clear-cell renal cell carcinoma. *Oncotarget* **2017**, *8*, 59777–59790. [\[CrossRef\]](#)
45. Xu, Z.; Liu, Y.; Yang, Y.; Wang, J.; Zhang, G.; Liu, Z.; Fu, H.; Wang, Z.; Liu, H.; Xu, J. High expression of Mucin13 associates with grimmer postoperative prognosis of patients with non-metastatic clear-cell renal cell carcinoma. *Oncotarget* **2017**, *8*, 7548–7558. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Chen, F.; Zhang, Y.; Şenbabaoğlu, Y.; Ciriello, G.; Yang, L.; Reznik, E.; Shuch, B.; Micevic, G.; De Velasco, G.; Shinbrot, E.; et al. Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma. *Cell Rep.* **2016**, *14*, 2476–2489. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Forbes, S.A.; Beare, D.; Boutselakis, H.; Bamford, S.; Bindal, N.; Tate, J.; Cole, C.G.; Ward, S.; Dawson, E.; Ponting, L.; et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **2017**, *45*, D777–D783. [\[CrossRef\]](#)
48. Ricketts, C.J.; De Cubas, A.A.; Fan, H.; Smith, C.C.; Lang, M.; Reznik, E.; Bowlby, R.; Gibb, E.A.; Akbani, R.; Beroukhi, R.; et al. The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Rep.* **2018**, *23*, 313–326. [\[CrossRef\]](#)
49. Girgin, C.; Tarhan, H.; Hekimgil, M.; Sezer, A.; Gürel, G. P53 mutations and other prognostic factors of renal cell carcinoma. *Urol. Int.* **2001**, *66*, 78–83. [\[CrossRef\]](#)
50. McDermott, D.F.; Huseni, M.A.; Atkins, M.B.; Motzer, R.J.; Rini, B.I.; Escudier, B.; Fong, L.; Joseph, R.W.; Pal, S.K.; Reeves, J.A.; et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat. Med.* **2018**, *24*, 749–757. [\[CrossRef\]](#)



51. Rini, B.I.; Powles, T.; Atkins, M.B.; Escudier, B.; McDermott, D.F.; Suarez, C.; Bracarda, S.; Stadler, W.M.; Donskov, F.; Lee, J.L.; et al. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): A multicentre, open-label, phase 3, randomised controlled trial. *Lancet* **2019**, *393*, 2404–2415. [\[CrossRef\]](#)
52. Loeb, L.A.; Bielas, J.H.; Beckman, R.A. Cancers exhibit a mutator phenotype: Clinical implications. *Cancer Res.* **2008**, *68*, 3551–3557. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [\[CrossRef\]](#)
54. Tsodikov, O.V.; Ivanov, D.; Orelli, B.; Staresinic, L.; Shoshani, I.; Oberman, R.; Schärer, O.D.; Wagner, G.; Ellenberger, T. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J.* **2007**, *26*, 4768–4776. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Gregg, S.Q.; Robinson, A.R.; Niedernhofer, L.J. Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA Repair* **2011**, *10*, 781–791. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Tufegdžić Vidaković, A.; Mitter, R.; Kelly, G.P.; Neumann, M.; Harreman, M.; Rodríguez-Martínez, M.; Herlihy, A.; Weems, J.C.; Boeing, S.; Encheva, V.; et al. Regulation of the RNAPII pool is integral to the DNA damage response. *Cell* **2020**, *180*, 1245–1261. [\[CrossRef\]](#)
57. Melis, J.P.M.; Luijten, M.; Mullenders, L.H.F.; van Steeg, H. The role of XPC: Implications in cancer and oxidative DNA damage. *Mutat. Res.* **2011**, *728*, 107–117. [\[CrossRef\]](#)
58. Apte, S.S. A disintegrin-like and metalloprotease (repolyisin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: Functions and mechanisms. *J. Biol. Chem.* **2009**, *284*, 31493–31497. [\[CrossRef\]](#)
59. The UniProt Consortium. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [\[CrossRef\]](#)
60. Matthews, L.; Gopinath, G.; Gillespie, M.; Caudy, M.; Croft, D.; de Bono, B.; Garapati, P.; Hemish, J.; Hermjakob, H.; Jassal, B.; et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **2009**, *37*, D619–D622. [\[CrossRef\]](#)
61. Adams, J.C.; Tucker, R.P. The thrombospondin type 1 repeat (TSR) superfamily: Diverse proteins with related roles in neuronal development. *Dev. Dyn.* **2000**, *218*, 280–299. [\[CrossRef\]](#)
62. Moremen, K.W.; Tiemeyer, M.; Nairn, A.V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 448–462. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Moch, H.; Cubilla, A.L.; Humphrey, P.A.; Reuter, V.E.; Ulbright, T.M. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.* **2016**, *70*, 93–105. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Singer, J.; Ruscheweyh, H.-J.; Hofmann, A.L.; Thurnherr, T.; Singer, F.; Toussaint, N.C.; Ng, C.K.Y.; Piscuoglio, S.; Beisel, C.; Christofori, G.; et al. NGS-pipe: A flexible, easily extendable and highly configurable framework for NGS analysis. *Bioinformatics* **2018**, *34*, 107–108. [\[CrossRef\]](#)
65. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Picard. Available online: <http://broadinstitute.github.io/picard/> (accessed on 31 August 2017).
69. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*. [\[CrossRef\]](#)
70. Hofmann, A.L.; Behr, J.; Singer, J.; Kuipers, J.; Beisel, C.; Schraml, P.; Moch, H.; Beerewinkel, N. Detailed simulation of cancer exome sequencing data reveals differences and common limitations of variant callers. *BMC Bioinform.* **2017**, *18*, 8. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Roth, A.; Ding, J.; Morin, R.; Crisan, A.; Ha, G.; Giuliany, R.; Bashashati, A.; Hirst, M.; Turashvili, G.; Oloumi, A.; et al. JointSNVMix: A probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* **2012**, *28*, 907–913. [\[CrossRef\]](#)
72. Cibulskis, K.; Lawrence, M.S.; Carter, S.L.; Sivachenko, A.; Jaffe, D.; Sougnez, C.; Gabriel, S.; Meyerson, M.; Lander, E.S.; Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **2013**, *31*, 213–219. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Kockan, C.; Hach, F.; Sarrafi, I.; Bell, R.H.; McConeghy, B.; Beja, K.; Haegert, A.; Wyatt, A.W.; Volik, S.V.; Chi, K.N.; et al. SiNVICT: Ultra-sensitive detection of single nucleotide variants and indels in circulating tumour DNA. *Bioinformatics* **2017**, *33*, 26–34. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Saunders, C.T.; Wong, W.S.W.; Swamy, S.; Becq, J.; Murray, L.J.; Cheetham, R.K. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **2012**, *28*, 1811–1817. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576. [\[CrossRef\]](#) [\[PubMed\]](#)

76. Ignatiadis, N.; Klaus, B.; Zaugg, J.B.; Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **2016**, *13*, 577–580. [\[CrossRef\]](#)
77. Lai, Z.; Markovets, A.; Ahdesmaki, M.; Chapman, B.; Hofmann, O.; McEwen, R.; Johnson, J.; Dougherty, B.; Barrett, J.C.; Dry, J. R. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **2016**, *44*, e108. [\[CrossRef\]](#)
78. Favero, F.; Joshi, T.; Marquard, A.M.; Birkbak, N.J.; Krzystanek, M.; Li, Q.; Szallasi, Z.; Eklund, A. C. Sequenza: Allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **2015**, *26*, 64–70. [\[CrossRef\]](#)
79. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [\[CrossRef\]](#)
80. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [\[CrossRef\]](#)
81. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Wang, J.; Vasaikar, S.; Shi, Z.; Greer, M.; Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **2017**, *45*, W130–W137. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Sturm, M.; Schroeder, C.; Bauer, P. SeqPurge: Highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinform.* **2016**, *17*, 208. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Smith, T.; Heger, A.; Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **2017**, *27*, 491–499. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Wong, S.Q.; Li, J.; Tan, A.Y.-C.; Vedururu, R.; Pang, J.-M.B.; Do, H.; Ellul, J.; Doig, K.; Bell, A.; MacArthur, G.A.; et al. CANCER 2015 Cohort. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med. Genom.* **2014**, *7*, 23. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Oh, E.; Choi, Y.-L.; Kwon, M.J.; Kim, R.N.; Kim, Y.J.; Song, J.-Y.; Jung, K.S.; Shin, Y.K. Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples. *PLoS ONE* **2015**, *10*, e0144162. [\[CrossRef\]](#) [\[PubMed\]](#)
87. Yost, S.E.; Smith, E.N.; Schwab, R.B.; Bao, L.; Jung, H.; Wang, X.; Voest, E.; Pierce, J.P.; Messer, K.; Parker, B.A.; et al. Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. *Nucleic Acids Res.* **2012**, *40*, e107. [\[CrossRef\]](#)
88. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **2016**, *32*, 292–294. [\[CrossRef\]](#)
89. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 1 April 2018).
90. Cingolani, P.; Patel, V.M.; Coon, M.; Nguyen, T.; Land, S.J.; Ruden, D.M.; Lu, X. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* **2012**, *3*, 35. [\[CrossRef\]](#)
91. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [\[CrossRef\]](#)
92. Moore, A.L.; Kuipers, J.; Singer, J.; Burcklen, E.; Schraml, P.; Beisel, C.; Moch, H.; Beerenwinkel, N. Intra-tumor heterogeneity and clonal exclusivity in renal cell carcinoma. *BioRxiv* **2018**. [\[CrossRef\]](#)
93. R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2018; Available online: <https://www.R-project.org/> (accessed on 31 October 2014).
94. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Tuszynski, J. caTools: Tools: Moving window statistics, GIF, Base64, ROC AUC, etc. R Package Version 1.17.1. 2014. Available online: <https://CRAN.R-project.org/package=caTools> (accessed on 1 April 2018).
97. Hadley Wickham, R.F.; Lionel Henry Müller, K. dplyr: A Grammar of Data Manipulation. R Package Version 0.7.4. 2017. Available online: <https://CRAN.R-project.org/package=dplyr> (accessed on 1 February 2018).
98. Wickham, H. ggplot2: Elegant Graphics for Data Analysis; Springer: New York, NY, USA, 2009.
99. Kassambara, A. ggpubr: “ggplot2” Based Publication Ready Plots. R Package Version 0.1.6. 2017. Available online: <https://CRAN.R-project.org/package=ggpubr> (accessed on 1 April 2018).
100. Gregory, R.; Warnes, B.B.; Lumley, T. gtools: Various R Programming Tools. R Package Version 350. 2015. Available online: <https://CRAN.R-project.org/package=gtools> (accessed on 1 November 2017).
101. Henningsen, A.; Toomet, O. maxLik: A package for maximum likelihood estimation in R. *Comput. Stat.* **2011**, *26*, 443–458. [\[CrossRef\]](#)
102. Kirill Müller, H.W. tibble: Simple Data Frames. R package version 1.4.2. 2018. Available online: <https://CRAN.R-project.org/package=tibble> (accessed on 1 April 2018).
103. Stefan Milton Bache, H.W. magrittr: A Forward-Pipe Operator for R. R package version 1.5. 2014. Available online: <https://CRAN.R-project.org/package=magrittr> (accessed on 1 April 2018).

- 
104. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Softw.* **2007**, *21*. [[CrossRef](#)]
  105. Neuwirth, E.; RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. 2014. Available online: <https://CRAN.R-project.org/package=RColorBrewer> (accessed on 1 April 2018).
  106. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849. [[CrossRef](#)]
  107. Therneau, T.M. A Package for Survival Analysis in S. Version 2.38. 2015. Available online: <https://CRAN.R-project.org/package=survival> (accessed on 1 April 2018).